

# Radial Distribution Studies Under Highly Constrained Conditions

Ronald G. Munro\*

National Bureau of Standards, Washington, D.C. 20234

September 5, 1979

The consequences of limited scattering data are considered for the determination of radial distribution functions. Such considerations are important, e.g., when substances are held at extreme pressure in a pressure vessel like the diamond anvil cell. By means of formal relations, alternatives to the direct Fourier inversion of the scattering data are considered, but it is found that they do not usefully circumvent the problems resulting from the truncation of data. Using an ideal set of data, five numerical procedures for inverting the data are compared as a function of the degree of data limitation. An extended-integral method is found to be the most reliable.

Key words: Diamond pressure cell; radial distribution functions; truncation error.

## 1. Introduction

The study of nonperiodic structures by means of radial distribution functions is well developed in the literature [1-4] for a variety of circumstances. However, the conditions now being encountered in the rapidly developing field of high pressure physics warrant a further investigation of the commonly used numerical procedures and a discussion of possible alternatives. These conditions occur in the study of condensed systems at large pressures using the diamond anvil pressure cell (DAC). To obtain large hydrostatic pressures, the cell body (the press) and a containing gasket must be made of materials (e.g. waspalloy or inconel) that are usually opaque to x-rays. Typically, the x-ray scattering angle ( $2\theta$ ) is  $< 15$  degrees. As a result, there is a natural preference for a fixed angle energy dispersive x-ray scattering technique. In this experimental method, a white beam of x-ray radiation is used with an energy sensitive detector. However, the useful energy range is typically [5,6] limited to  $10 \text{ keV} \leq E \leq 40 \text{ keV}$ . The lower limit is determined by the strong absorption in the diamond windows, and the upper limit is determined by the diffraction efficiency which is proportional to  $E^{-2}$ . The result for the DAC is a set of scattering data that is more limited than is commonly found for x-ray experiments at one atmosphere.

The calculation of a radial distribution function (RDF) from x-ray scattering data requires the evaluation of a Fourier sine transform. When a Fourier transform defined on an infinite interval is evaluated by means of experimentally determined values of the integrand, there are three principal sources of error contributing to the computed value of the transform: the uncertainty in the integrand's value at a point, the number of points at which the value of the integrand is known (the information density), and the actual interval over which the transform is computed (the information range). The first two error sources are intimately related to the accuracy and resolution of the experimental measurements. The third error, the termination error, occurs whenever a finite interval is used instead of the infinite interval.

Under the restrictions of the DAC, it is especially of interest to know the relative effectiveness of various techniques for minimizing the effects of limited data and to have an indication of the circumstances under which the methods fail to be reliable. Although various computational techniques have been discussed individually in the literature, a direct comparison of their performances under controlled conditions of variable difficulty has not been given. It is the purpose of this work to present such a comparison with a particular interest in the conditions of the DAC.

The RDF is of special interest in liquid and amorphous systems which are lacking a long-ranged periodic structure.

<sup>1</sup> Figures in brackets indicate literature references at the end of this paper.

\*Center for Materials Science, National Measurement Laboratory.

For such systems, we consider the correlation of the occurrence of a particle at a distance  $r$  from a reference particle which is assumed to be at the origin. The resolution of structural information decreases as the distance from the reference particle increases. Consequently, the likelihood of finding another particle at large distance  $r$  is simply proportional to the average sample density. Let  $\varrho(r)$  be the correlation density function and set  $\varrho(r) = \bar{\varrho}g(r)$ . The quantity  $g(r)$  is the RDF and  $\bar{\varrho}$  is the average density. The normalization of  $g(r)$  is such that  $\lim_{r \rightarrow \infty} g(r) = 1$ . Since no other particle will occur at the origin given that the reference particle is there, we also have  $g(0) = 0$ .

The differential RDF which is defined as

$$D(r) = 4\pi r [\varrho(r) - \bar{\varrho}] \quad (1)$$

is a more convenient function for discussion. It is well known that  $D(r)$  is related to the observable scattering intensity data  $F(s)$  by Fourier sine integrals of the forms

$$D(r) = \frac{2}{\pi} \int_0^\infty F(s) \sin(rs) ds \quad (2)$$

$$F(s) = \int_0^\infty D(r) \sin(rs) dr \quad (3)$$

where  $s = 4\pi \sin\theta/\lambda$  is the scattering variable for scattering angle  $2\theta$  and x-ray wavelength  $\lambda$ . In more common notation,  $F(s) = s i(s)$  where  $i(s)$  is the reduced scattered intensity. Experimental observations are made for only a finite number of values of  $s$  so that  $F(s)$  is known only for  $s_{\min} \leq s \leq s_{\max}$ . One result [7] of using the interval  $(s_{\min}, s_{\max})$  instead of  $(0, \infty)$  in eq (2) is that  $D(r)$  acquires a modulation with frequency components in  $r$ -space of the order of  $1/s_{\min}$  and  $1/s_{\max}$ . Furthermore, the locations and widths of true extrema are shifted by amounts that depend on the degree of truncation. Common smoothing techniques do not correct these problems, as will be seen, e.g., in section 4.

A number of procedures have been used to treat termination problems. Some comments on the utility of these techniques and also the problems generated by them are made in the following sections. Let us summarize some of the procedures. For  $s < s_{\min}$  when  $s_{\min}$  is small, the use of a formula to extrapolate the intensity data to  $s = 0$  has been suggested [8], and this is usually found to be adequate. For large  $s$ , a convergence factor of the form  $e^{-as^2}$  is often added to the integrand in eq (2) to act as a smoothing function. Other modification functions [2] have also been considered. Since spurious features are directly dependent on the integration cut off values, attempts have been made to identify the spurious features by considering a sequence [9,10] of limits.

The thought is that the true peaks in the  $D(r)$  curve should be relatively stationary, so peaks showing significant shifts with limit changes can be deleted. Another possible approach is to smooth the computed curve by inspection and then back-transform [11] using eq (3) to see the effect. Yet another approach [12] couples the smoothing operation to the data reduction stage of the problem where  $F(s)$  is determined from the raw data by applying several experimentally required corrections. The objective of each of these techniques is to obtain a smooth curve with well defined peaks.

However, dissatisfaction with the treatment of the termination problems is still expressed in the literature for several reasons. If the scattering data are smoothed to meet the expectations of the observer, then the quantitative result for the RDF is coupled to the observer's qualitative acceptance of details in the scattering data. When smoothing or modification functions are used, it is not clear how one can unambiguously demonstrate that the procedures have introduced less error than they have removed. For example, when a convergence factor  $e^{-as^2}$  is added to the integrand of eq (2), the number, locations, widths, and amplitudes of the peaks in the resulting RDF are dependent on the strength of the convergence factor. Even in the procedure of back-transforming the RDF to show selfconsistency with the final scattering data, there is a limitation on the attainable accuracy because the integration step size is constrained by the experimental resolution. Finally, there is the problem of automating the analysis which is relevant not only to the convenience of the procedure but also to the reproducibility of the result.

In addition to the location of the shells of neighboring particles, it should also be possible to determine the number of particles contained in each neighboring shell. This number should be determined by the area under the peak of the RDF curve. However, termination error and the actual or effective introduction of smoothing functions produce an  $r$ -dependent error which is rather significant in its effect on the area. Recognizing this problem, at least one method [12] of analysis uses the number of particles in a shell as a parameter whose value can be fixed or treated as adjustable.

In the following, we (1) formalize the statement of the problem of limited data, (2) use the formal results to examine the nature of the error, (3) indicate how some of the techniques mentioned above are related to the formal relations, (4) consider alternative methods of surmounting the truncation error, and (5) illustrate, compare, and discuss the applications and limitations of the various objective methods.

In general, the results show that the more common procedures do not perform well when the data set is highly limited. The most reliable results appear to be obtained by the extended-integral method of Hansen [13-15] et al.

## 2. Formal relations

In this section, a number of formal relations are examined in some detail. From these, we obtain a clear indication of the problems associated with the calculation of a radial distribution function by means of a truncated Fourier sine transform, and we are given an opportunity to evaluate the potential utility of alternative approaches to the RDF.

### 2.1 Information properties

The error in  $D(r)$  produced by truncating the transform integral is a function of  $r$ , and in fact, the nature of the error is also a function of  $r$ . We have devised a compact schematic means of illustrating the changing nature of the problem by converting the transform integral on an infinite interval to an integral on the continuous interval  $(0, 2\pi)$ . Define the integration variable  $\theta$  by the equation  $rs = \tan(\theta/4)$ . Then,

$$D(r) = \frac{1}{2\pi r} \int_0^{2\pi} F[s(r, \theta)] \sin[\tan(\theta/4)] \sec^2(\theta/4) d\theta \quad (4)$$

If we interpret the integration variable  $\theta$  as an angle, then we can define a plane, which we shall call the information plane, that is described by the polar coordinates  $(r, \theta)$ . In this plane, the transformed integral is a line integral on a closed curve (circle).

Suppose that  $F(s)$  is known for  $s \leq 15$  with a resolution of  $\delta s = 0.05$ . Then, for each fixed value of  $r$ , there is a finite set of values of  $\theta$  at which the integrand of eq (4) is known. By using a plane described by polar coordinates  $(r, \theta)$ , the distribution of the discrete values of  $\theta$  can be shown as points on circles of constant  $r$ . This is done in figure 1 for the case mentioned. It should be noted that this construction does not depend on the particular substance described by  $F(s)$ . It provides an investigation of information density and range.

Examination of eq (4) reveals that the integrand oscillates with a rapidly increasing frequency as  $\theta$  increases. Thus, it is most desirable to have a large information density in the fourth quadrant of figure 1. For  $r = 0.1$ , there are no points in the fourth quadrant, and it must be expected that  $D(r)$  for  $r \sim 0.1$  is not reliably calculated for  $S_{\max} = 15$ ,  $\delta s = 0.05$ . For  $r = 1$ , there is a nearly uniform density over the first three quadrants with a somewhat larger density in the fourth quadrant. In this case, there are information samplings from all regions of the plane, but there is a terminal gap of about 15 degrees in the fourth quadrant, and the density of points is not large anywhere. Thus,  $r = 1$  is a borderline case. For those applications requiring high accuracy,  $D(r)$  for  $r \sim 1$  probably contains significant error. The case  $r = 15$  shows a problem of the opposite extreme. Only three points are not

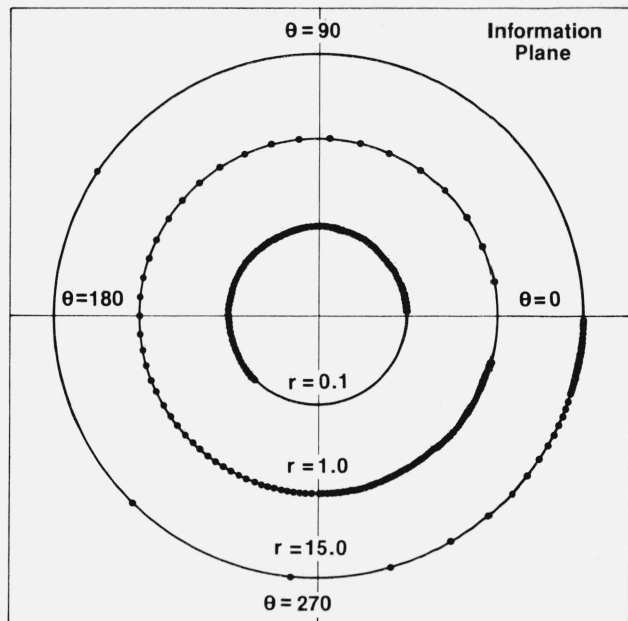


FIGURE 1. The information plane showing the points in  $r, \theta$  space at which the intensity function for  $F(s)$  is known when  $s \leq 15$  and  $\delta s = 0.05$ .

The solid bands represent closely spaced points.

in the fourth quadrant, and the terminal gap is only about 1 degree. However, integration over the first three quadrants is quite questionable, and the density of points in the fourth quadrant is rather sparse over about 2/3 of the range. This problem is specifically related to the resolution  $\delta s$ .

In general, the most significant consideration for small  $r$  ( $r \sim 1$ ) is the information range, while for larger  $r$ , the information density becomes increasingly important. The information plane shows these features in a clear and convenient way.

### 2.2 Eigenfunction expansion

Another important characteristic of the Fourier sine transform related to the foregoing discussion is that the integral is illconditioned. For present purposes, we can say roughly that the integral is illconditioned if an accurate evaluation of the integral in any region of  $r$ -space requires information samplings from all regions of  $s$ -(or  $\theta$ -) space. The impact of this condition can be seen by means of an eigenfunction expansion in terms of the eigenfunctions of the integral operator.

The eigenfunctions  $\psi$  of the integral transform operator are defined by the equation

$$\int_0^\infty \psi_\lambda(s) \sin(rs) ds = \lambda \psi_\lambda(r). \quad (5)$$

For the present discussion, it is sufficient to know that a complete mutually orthogonal set of functions  $\psi_\lambda$  exists [16]. These functions are oscillatory and cannot be normalized, but they can be chosen such that

$$\int_0^\infty \psi_\lambda(\chi) \psi_\mu(\chi) d\chi = \delta(\lambda - \mu). \quad (6)$$

Consequently, we can write  $D(r)$  as an expansion in the  $\psi_\lambda$ .

$$D(r) = \int_0^\infty A(\lambda) \psi_\lambda(r) d\lambda \quad (7)$$

where

$$A(\lambda) = \lambda^{-1} \int_0^\infty F(s) \psi_\lambda(s) ds. \quad (8)$$

We see immediately that every coefficient in the expansion suffers the effects of truncation error. Furthermore, there is no natural cutoff for the number of coefficients retained in the expansion [16]. Consequently the eigenfunction expansion method cannot be used reliably unless there is a large information range with a high information density.

### 2.3 Direct correlation function

In the distribution function formulation of the theory of liquids and amorphous materials, the introduction of a direct correlation function  $c(r)$  has been useful [17]. This function has a shorter range than does the RDF, and it basically measures only the correlation of a reference particle with the particles that are nearest to it. The definition of  $c(r)$  is given in terms of the Ornstein-Zernike equation.

$$D(r_{12}) = 4\pi \bar{\rho} r_{12} c(r_{12}) + \bar{\rho} r_{12} \int_0^\infty \frac{D(R_{13})}{R_{13}} c(R_{23}) dR_3 \quad (9)$$

where we have momentarily written  $r_{12}$  for  $r$  and  $R_{ij} = R_i - R_j$ . Starting from eq (9), we can obtain an expression for  $c(r)$  in terms of the scattering function  $F(s)$ . Using the properties of the Fourier sine transform, we obtain

$$4\pi \bar{\rho} r c(r) = \frac{2}{\pi} \int_0^\infty \frac{s F(s) \sin(rs) ds}{s + F(s)} \quad (10)$$

For large values of  $s$ , when  $F(s) \ll s$ , and the integrand in eq (10) is approximately  $F(s) \sin(rs)$ . Hence, the evaluation of the direct correlation function by means of scattering data suffers truncation error in the same manner as does the full RDF. Therefore, the termination problem of the RDF cannot be circumvented by this approach. The fact that the problem persists undiminished for the function  $c(r)$ , even though it is simpler in structure than the RDF, is a further

illustration of the illconditioned nature of the sine transform. Before leaving this point, we consider another proposition that encounters some difficulty for essentially the same reason.

### 2.4 Iterative solutions

The effects of termination error would be quite unimportant if the scattering function  $F(s)$  were negligibly small at and beyond the upper limit of the numerical integration. A procedure that is often used to simulate this condition is to insert a convergence factor into the integrand. Then, instead of computing  $D(r)$ , a quantity  $Q(r)$  is obtained. Let  $\Theta(s)$  be the convergence factor. Then  $Q(r)$  is given by the following integral.

$$Q(r) = \frac{2}{\pi} \int_0^\infty F(s) \Theta(s) \sin(rs) ds \quad (11)$$

In equation (11),  $\Theta(s)$  is any function which actually or effectively restricts the integration to the interval on which  $F(s)$  is known. Equation (2) and (11) are Fourier transform relations that differ only by the additional factor  $\Theta(s)$  in eq (11), and hence,  $Q(r)$  and  $D(r)$  are related by a convolution relation. Defining the cosine Fourier transform  $\theta(x)$  by the relations

$$\theta(x) = \frac{2}{\pi} \int_0^\infty \tilde{\Theta}(s) \cos(xs) ds \quad (12)$$

$$\tilde{\Theta}(s) = \int_0^\infty \theta(x) \cos(sx) dx \quad (13)$$

We obtain the equation expressing  $Q(r)$  in terms of  $D(r)$ .

$$Q(r) = \int_0^\infty \frac{1}{2} [D(r-x) + D(r+x)] \theta(x) dx \quad (14)$$

A few remarks about the functions  $\tilde{\Theta}(s)$  and  $\theta(x)$  are useful at this point. If  $\theta(x)$  is chosen to be a delta function,  $\theta(x) = \delta(x)$ , then  $Q(r) = D(r)$ . This case corresponds to knowing  $F(s)$  for all  $s$  so that  $\tilde{\Theta}(s) = 1$ . Given only a limited range for  $s$ , the more closely  $\theta(x)$  can be chosen to resemble a delta function, the more closely  $Q(r)$  approximates  $D(r)$ . Although the ideal case for  $\Theta(s)$  is the constant unity, in practice  $\tilde{\Theta}(s)$  must fall to a sufficiently small value at  $s = s_{\max}$  so that the range  $s > s_{\max}$  makes a negligible contribution. One choice for  $\theta(x)$  which approximately produces both desired features is the function

$$\theta(x) = (\pi a)^{-1/2} \exp(-x^2/4a)$$

which yields the often used function

$$\Theta(s) = e^{-as^2}$$



Furthermore, it should be clear that a sequence of expressions for  $\theta(x)$  which tends toward  $\delta(x)$  should generate a sequence of  $Q(r)$  which tends toward  $D(r)$ . For the Gaussian  $\theta(x)$  mentioned above, let  $a = b^2/s_{\max}^2$ . Then, for a fixed dimensionless parameter  $b$ , one way to obtain a sequence of  $Q(r)$  is to take a sequence of cutoff values  $s_{\max}$ . The use of this procedure is mentioned in the Introduction. Here, we see that the general approach is to consider a sequence of functions  $\theta(x)$ .

Equation (14) can be used to obtain  $D(r)$  by iteration. Let  $W(x)$  be a measure of the amount by which  $\theta(x)$  differs from  $\delta(x)$ .

Formally, we can write

$$\theta(x) = \delta(x) + W(x) \quad (15)$$

Substituting (15) into (14), we find

$$D(r) = Q(r) - \int_0^\infty (1/2) [D(r-x) + D(r+x)] W(x) dx. \quad (16)$$

Equation (16) provides an expression for  $D(r)$  which can be used for  $D(r-x)$  and  $D(r+x)$ . Substitution of these latter expressions into the integrand of the right hand side of eq (16) yields a new expression for  $D(r)$ . By successive substitutions, one obtains from equation (16) the result

$$D(r) = Q(r) + \sum_{K=1}^{\infty} (-1/2)^K \sum_{L=0}^K C(K,L) \int_0^\infty W(X_1) dX_1 \int_0^\infty W(X_K) dX_K Q(r-X_1 - \dots - X_{K-L} + X_{K-L+1} + \dots + X_K). \quad (17)$$

where  $C(K,L) = K!/L!(K-L)!$ . Equation (17) is an exact relation for  $D(r)$ . However, from the form of eq (17) it would seem that the evaluation of  $D(r)$  by means of this series would present some difficulty. One problem is the explicit determination of the function  $W(x)$ . Another is that convergence of the series appears to depend partly on a cancellation effect among the terms in the summation on  $L$  for each fixed  $K$  in eq (17). The problem can be made quite clear by considering the approximation obtained by terminating the series after the first iteration. This gives

$$D(r) \cong Q(r) - 1/2 \int_0^\infty [Q(r-x) + Q(r+x)] W(x) dx. \quad (18)$$

This approximation is good if  $W(x)$  is similar to the function  $W_1(x)$  which would be required to make eq (18) an exact equality. The function  $W_1(x)$  is easily shown to be

$$W_1(x) = \delta(x) - \frac{2}{\pi} \int_0^\infty \frac{\cos(xs)}{\Theta(s)} ds. \quad (19)$$

from which it is seen that  $W(x)$  does not readily produce convergence in eq (17).

The problem of choosing the function  $W(x)$  can be avoided by an alternative iterative procedure. The estimate  $Q(r)$  is obtained from eq (2) by inserting a convergence factor into the integrand of the transform integral. The error of the estimate is

$$D(r) - Q(r) = \frac{2}{\pi} \int_0^\infty F(s) [1-\Theta(s)] \sin(rs) ds. \quad (20)$$

The integral in this equation cannot be evaluated, but an estimate of the error can again be obtained by inserting a factor  $\Theta(s)$  into the integrand of eq (20). After each estimate, the procedure can be repeated. Let  $Q_K(r)$  be a generalization of  $Q(r)$  such that we have

$$Q_K(r) = \frac{2}{\pi} \int_0^\infty F(s) [1-\Theta(s)]^K \Theta(s) \sin(rs) ds \quad (21)$$

Then, it is easy to see that

$$D(r) = \sum_{K=0}^{\infty} Q_K(r). \quad (22)$$

It is also easy to see that the series (22) does not converge rapidly. The product  $[1-\Theta(s)]^K \Theta(s)$  is never negligible for all values of  $s$  for a function like  $\Theta(s) = e^{-as^2}$ . Suppose however, that the product is sharply peaked so that only a small range of  $s$  about  $s = s(K)$  is important. Then,  $Q_K(r) \cong \text{constant} \times \sin[rs(K)]$ , and even in this case, a long ranged oscillatory contribution would be obtained.

The problems here and in the preceding subsections are illustrations of the illconditioning of the sine transform and indicate the importance of the information distribution. In other words, all of the approaches to the evaluation of  $D(r)$  encounter comparable difficulties because the basic problem is inherent to the transform. From the discussion in section 2, it is clearly desirable to consider the evaluation of the integral in eq (2) directly rather than proceeding indirectly or iteratively. In the following section, some of the methods used to reduce the effect of limited information are compared.

### 3. Practical procedures

A number of techniques to minimize the errors produced by a limited amount of information are currently in use. These methods can be classified into two types, those which assume a detailed crystalline-type model of the structure of the system and those which do not. The present work is concerned with methods of the second type. These procedures involve reasonable assumptions about the function  $D(r)$  or about the behavior of  $F(s)$  at large values of  $s$ .

The use of a convergence factor and the method of back-transforming to selfconsistency have already been mentioned. Both of these techniques have desirable features. The modulation of the RDF produced by an abrupt termination of the integration interval is reduced by the convergence factor, and the requirement of selfconsistency is a means of increasing the information content of the data. A variation of these two techniques, is provided by an indirect selfconsistent method. If we let  $D(r)$  and  $Q(r)$  be given respectively by eq (2) and (11), we can write in general

$$D(r) = Q(r) + E(r) \quad (23)$$

where  $E(r)$  is a function representing the error. The approximation  $E(r) = 0$  yields the convergence factor method. The function  $Q(r)$  is determined primarily by the most accurate and reliable portion of the scattering data because of the factor  $\Theta(s)$  in expression (11). The correction term  $E(r)$  is then determined primarily by the data at larger  $s$  values. Experimentally, the uncertainty in the scattering function  $F(s)$  increases as  $s$  increases, and hence it is the function  $E(r)$  that would benefit most significantly by the additional condition of selfconsistency. However, when the condition of selfconsistency is applied directly to  $D(r)$ , the elements of the scattering data are treated equally. Consequently, the suggestion is that  $Q(r)$  be taken as a first estimate of  $D(r)$  and that a selfconsistent correction term  $E(r)$  then be obtained to complete the evaluation of  $D(r)$ . By splitting the evaluation of  $D(r)$  into two parts, the most reliable data can be emphasized preferentially. Also, the correction term is expected to be a smaller contribution to  $D(r)$  than is  $Q(r)$ . As a result, small errors in  $E(r)$  might be expected to be of second order smallness in  $D(r)$ .

From eq (2), (11), and (23) we have

$$E(r) = \frac{2}{\pi} \int_0^\infty F(s) [1 - \Theta(s)] \sin(rs) ds \quad (24)$$

As a first estimate of  $E(r)$ , we can use the quantity  $Q_1(r)$  given by eq (21). The well known numerical techniques or back-transforming can then be used to produce selfconsistency. Furthermore, we can use the convolution relation

$$Q_1(r) = 1/2 \int_0^\infty [E(r-x) + E(r+x)] \Theta(x) dx \quad (25)$$

to examine selfconsistency since eq (21) and (25) provide two different evaluations of the function  $Q_1(r)$ .

The last technique that we consider is the extended-integral method of Hansen [13-15] et al. This procedure uses the observation that the quantity  $F(s)$  at large  $s$  is determined primarily by the structure at short distances. Hansen [13-15] et al. have found it most useful to assume a Gaussian model for the distribution of near-neighbor atoms and

to assume a uniform correlation density at distances beyond the near-neighbors. This model is given by the following relations.

$$D_C(r) + 4\pi r \bar{Q} = \sum_K N_K (2\pi\sigma_K^2)^{-1/2} (1/R_K) \exp[-(r-R_K)^2/2\sigma_K^2] \quad (26)$$

for  $r < R_i$ , and  $D_C(r) = 0$  for  $r > R_i$ , where  $R_i$  is a measure of the near-neighbor range.

Correspondingly, the function  $F_C(s)$  is

$$F_C(s) = \sum_K N_K \frac{\sin(R_K s)}{R_K} \exp[-1/2\sigma_K^2 s^2] + \frac{4\pi\bar{Q}}{s^2} [s R_i \cos(R_i s) - \sin(R_i s)] \exp[-1/2\sigma_i^2 s^2]. \quad (27)$$

The number  $\sigma_i$  has been introduced into eq (27) for the purpose of smoothing the transition between the two different regions of the model. For most cases, the last term of eq (27) is negligible for large values of  $s$ .

Next, the analytic model  $F_C(s)$  and the observed  $F(s)$  are transformed numerically on the observed set of  $s$ -values. This results in the same truncation errors being produced in both transforms. The transformed model can then be fit to the transform of the observed data by a least squared error method. The emphasis in the fitting procedure is given to the features in the near-neighbor range,  $r < R_i$ . Then, the assumption that  $F(s)$  at large  $s$  is determined by the structure at short distances allows the observed range of  $s$  to be extended by setting  $F(s) = F_C(s)$  for  $s > s_{\max}$ . In this manner, the effective information content of the data is significantly increased, and  $D(r)$  is evaluated by an untruncated integral.

## 4. Discussion

The evaluation of the RDF by means of the unmodified truncated integral, the convergence factor method, selfconsistency, indirect selfconsistency, and the extended-integral method represent the procedures which do not assume prior knowledge of the structure of the system. To compare and determine the effectiveness and limitations of these techniques, it is important that we use an example for which the source of error is known to be only that a finite set of scattering data is used. Such an example is not attainable experimentally. However, since the procedures under discussion are not dependent on a physical structure, a sufficient example is easily constructed. The following equations provide the example used here. The exact differential RDF is assumed to be

$$D(r) = \sum_K C_K \frac{B_K}{B_K^2 + (A_K - r)^2} - \frac{B_K}{B_K^2 + (A_K + r)^2} \quad (28)$$

for which the exact scattering function [18] is

$$F(s) = \pi \sum_K C_K e^{-B_K s^2} \sin(A_K s). \quad (29)$$

In figure 2, the scattering function  $F(s)$  is shown for the particular set of parameters ( $A_K$ ,  $B_K$ ,  $C_K$ ) used for the example. For real systems, the differential RDF would be zero for all distances smaller than the breadth of an atom, and consequently there would be a greater richness in high frequency Fourier components than in the selected example. This aspect of the problem has already been discussed in detail by Mountain [19] and need not be repeated here. Also indicated in the figure are three data ranges which are labeled minor, modest, and severe. These three ranges are used to examine the various procedures as a function of the degree of termination. In the minor case, very little information is lost by the truncation because the magnitude of  $F(s)$  is quite small for all  $s > 15$ . Consequently, all methods are expected to be adequate for this case. In the other extreme, the severe range involves a very significant loss of information. Common experimental situations fall between these cases and are represented by the modest range, but experiments with diamond anvil pressure cell constraints can fall within the severe to modest range. In each case, it is assumed that the data are known with a resolution  $\delta s = 0.05$ .

Two of the calculation procedures use a requirement of selfconsistency. In these cases, the computations continue until selfconsistency is obtained in both of the senses mentioned previously, i.e. by back-transforming and by means of the convolution relations. As an example, figure 3 shows the selfconsistency obtained by the convolution relation for the modest range.

For the extended-integral method, the  $D(r)$  fit at short distances and the smoothness of the extension of  $F(s)$  are of interest, and these results are illustrated by figures 4 and 5.

The functions  $D(r)$  computed by each of the procedures in each of the s-ranges are shown in figures 6-8. In these figures, the solid curve is the exact  $D(r)$  given by eq (28), and the plotted points are values calculated by the various methods.

It is not surprising that all of the procedures perform well in the case of minor truncation error as is shown in figure 6. Only a very small amount of information is not contained in the  $F(s)$  data, and even the truncated integral yields not more than a minor error at small  $r$ .

The fidelity of the computed RDF values to the exact RDF curve decreases as the data range is decreased. In figures 7 and 8, the most obvious effect is the reduction of the peak amplitudes of the computed  $D(r)$  curves. This has serious consequences for the evaluation of the number of particles in a neighboring shell since this number is determined

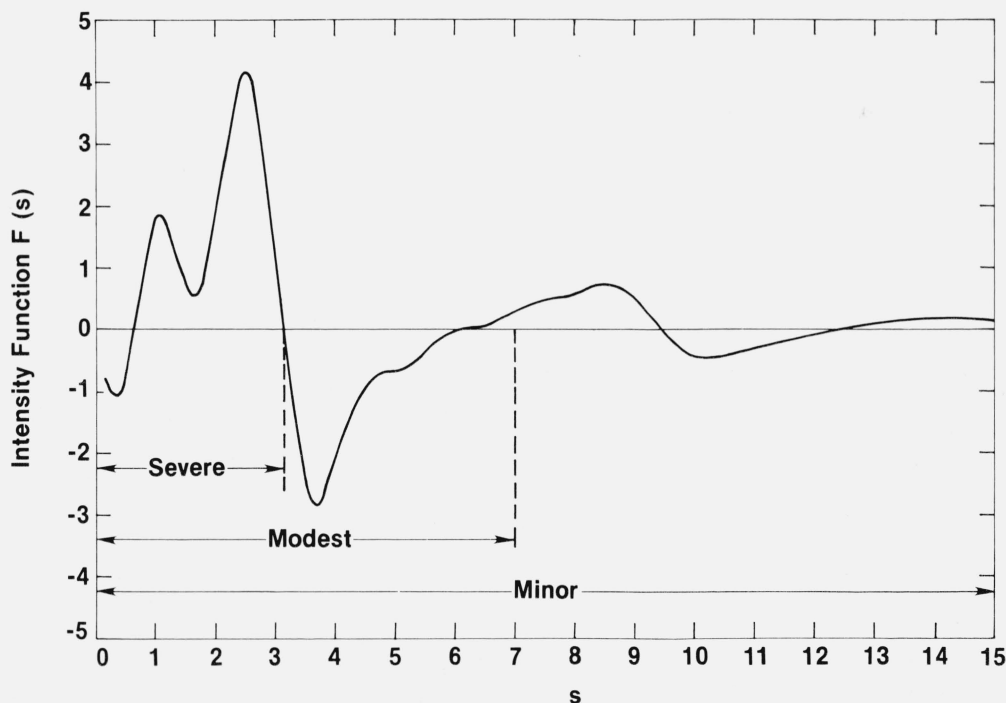


FIGURE 2. The intensity function  $F(s)$  given by eq (29) with  $(A_K, B_K, C_K) = (1.0, 0.2, 1.0), (2.0, 0.3, -1.0), (3.0, 0.4, 1.0), (4.0, 0.5, -1.0)$ .

The three truncation ranges considered in this paper are indicated by the labels severe, modest, and minor.

by the area under the RDF peak. In any problem involving an unusually limited data range, there appears to be little expectation of finding reliably the number of particles in a neighboring shell.

The location of the extrema are produced quite well by all methods in the minor and modest cases, and the spurious oscillations from the truncation are smoothed. However, in the severe data case, none of the methods is especially good. According to figure 8, the extended-integral method has the best performance, and it is the only method which removes the spurious extremum at  $r=5.5$ . Significant errors are still found for the locations of the first peak ( $r=0.82$  computed versus  $r=1.0$  exact) and the second minimum ( $r=4.26$  computed versus  $r=4.0$ ).

In figure 2, it is seen that the severe data range involves the loss of a significant feature in the  $F(s)$  curve. The data

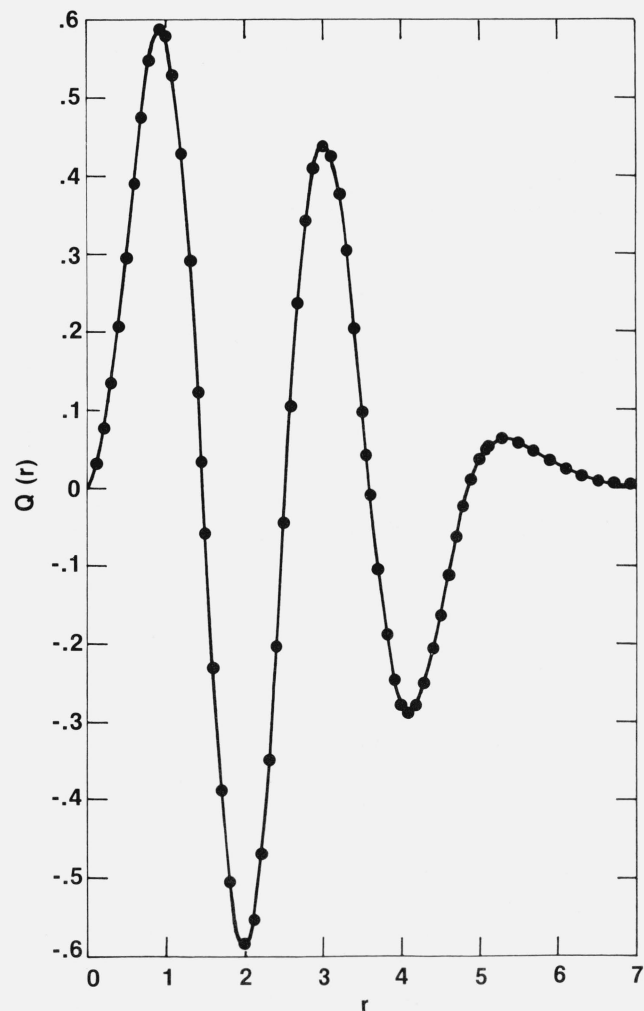


FIGURE 3. An illustration of the typical selfconsistency obtained in the selfconsistent calculation procedures.

The solid curve is given by a direct calculation of the function  $Q(r)$ . The plotted points are determined by means of the convolution relation eq (14).

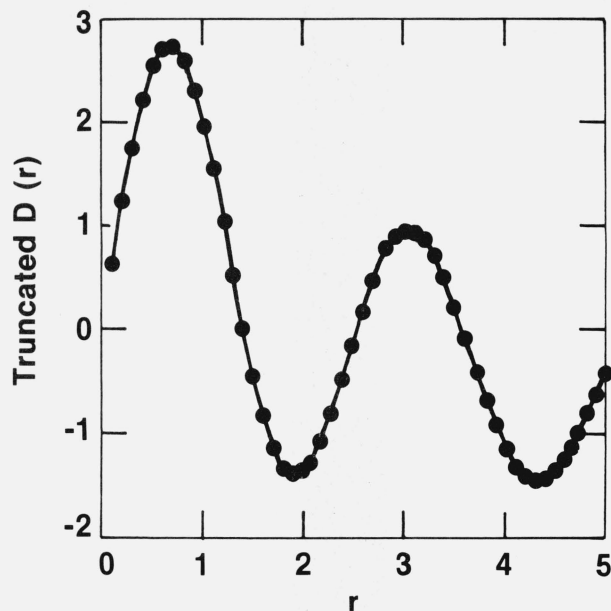


FIGURE 4. An illustration of the fit (plotted points) to the truncated integral (solid curve) obtained in the first step of the extended-integral calculation.

The corresponding extension of the intensity function is shown in figure 5.

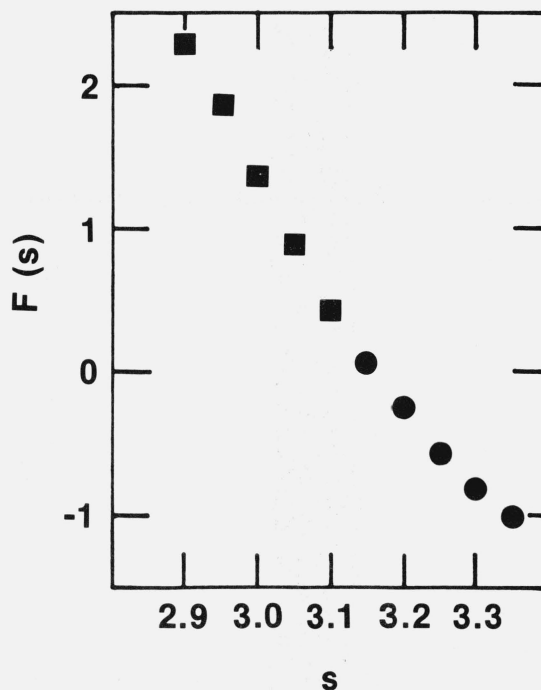


FIGURE 5. An illustration of the smooth continuity of the extension (circular points) of the given  $F(s)$  data (square points) achieved by the extended-integral method.

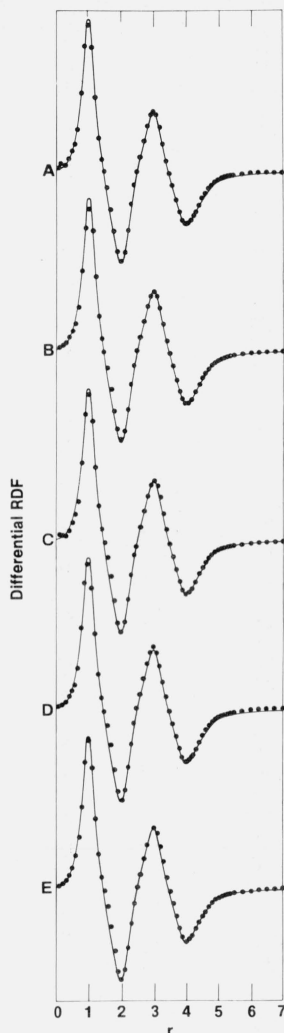


FIGURE 6. Differential RDF for the minor truncation of data in figure 1. Solid curves are exact.

Plotted points are computed by the methods: (A) simple truncation (B) convergence factor (C) direct selfconsistency (D) indirect selfconsistency (E) extended-integral.

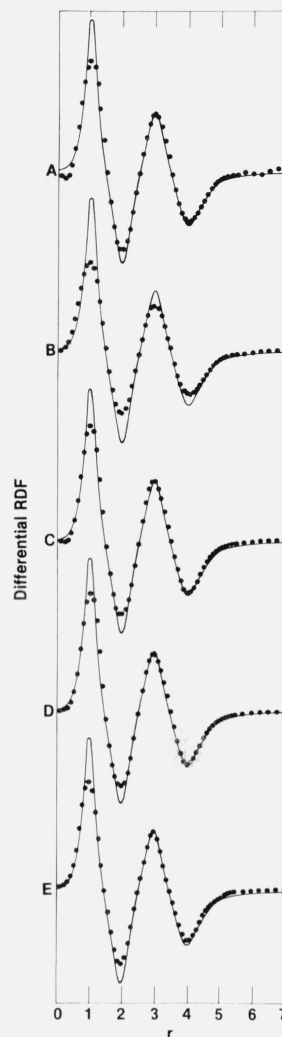


FIGURE 7. Differential RDF for the modest truncation of data in figure 1.

Solid curves are exact. Plotted points are computed by the methods: (A) simple truncation (B) convergence factor (C) direct selfconsistency (D) indirect selfconsistency (E) extended-integral.

lost by this truncation appears to be more structural than the data commonly found in this range of  $s$  for glasses and liquids. Often, the observed data nearly describe a damped sinusoid. Consequently, the severe data range of figure 2 represents a worse case than will be found in experiments with the diamond anvil cell. Furthermore, since the extended-integral method yields a damped sinusoidal extension of  $F(s)$ , the expected regularity of actual experimental data should make the extended-integral method even more effective.

With the DAC it is possible to extend the range of the scattering variable  $s=4\pi\sin\theta/\lambda$  by using special design features which permit access to angles  $2\theta>15$  degrees. The

challenge to the experimentalist is to produce such designs that do not sacrifice the high pressure hydrostatic capabilities of the cell. The results of the present work indicate that maximizing the  $s$ -range obtainable in a DAC should be considered a requirement for reliable RDF determinations.

The extended-integral method appears to be the preferred method of analysis. In section 2, it is seen that alternative approaches such as the direct correlation function or the iterative techniques do not alleviate the problems produced by limited data, and as a result, it is found to be better to work directly with the Fourier inversion of the scattering intensity data,  $F(s)$ . Whenever the data range has only a small degree of truncation, most of the common computa-



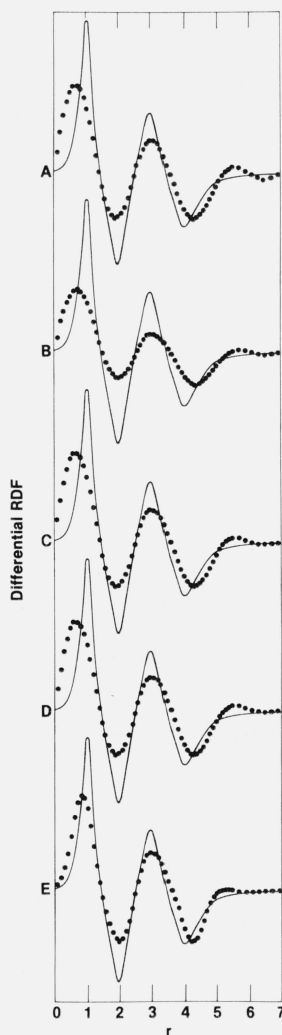


FIGURE 8. Differential RDF for the severe truncation of data in figure 1.

Solid curves are exact. Plotted points are computed by the methods: (A) simple truncation (B) convergence factor (C) direct selfconsistency (D) indirect selfconsistency (E) extended-integral.

tional procedures will determine the locations of the neighboring shells quite well. In such cases, the errors introduced through data correction and reduction procedures will be much more significant than the termination error. As the degree of truncation increases, the accuracy of the extended integral method deteriorates the least, and for this reason, it is the method which should be used in the RDF analysis when the conditions of the diamond anvil pressure cell prevail.

## 5. References

- [1] Waser, J., and Schomaker, V., *Rev. Mod. Phys.* **25**, 671 (1953).
- [2] Kruh, R. F., *Chem. Rev.* **62**, 319 (1962).
- [3] Paalman, H. H., and Pings, C. J., *Rev. Mod. Phys.* **35**, 389 (1963).
- [4] Klug, H. P., and Alexander, L. E., *X-ray Diffraction Procedures* (Wiley, 1974).
- [5] Shimomura, O., *Bull. Japan. Crystall. Soc.* **21**, 94 (1979).
- [6] Bordas, J., Glazer, A. M., Howard, C. J., and Bourdillon, A. J., *Phil. Mag.* **35**, 311 (1977).
- [7] Bragg W. L., and West, J., *Phil. Mag.* **10**, 823 (1930).
- [8] Mikolaj, P. G., and Pings, C. J., *J. Chem. Phys.* **46**, 1401 (1967).
- [9] Pings, C. J., and Paalman, H. H., *Mol. Phys.* **5**, 531 (1962).
- [10] Stirpe S., and Tompson, C. W., *J. Chem. Phys.* **36**, 392 (1962).
- [11] Kaplow, R., Strong, S. L., and Averbach, B. L., *Phys. Rev.* **138**, 1336 (1965).
- [12] Konnert, J. H., and Karle, J., *Acta Cryst.* **A29**, 702 (1973).
- [13] Hansen, F. Y., Knudsen, T. S., and Carneiro, K., *J. Chem. Phys.* **62**, 1556 (1975).
- [14] Hansen F. Y., and Carneiro, K., *Nuc. Instrum. Meth.* **143**, 569 (1977).
- [15] Hansen, F. Y., *Comput. Phys. Commun.* **15**, 401 (1978).
- [16] McWhirter J. G., and Pike, E. R., *J. Phys.* **A11**, 1729 (1978).
- [17] McQuarrie, D. A., *Statistical Mechanics*, (Harper and Row 1976), p. 268.
- [18] Erdelyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G., eds., *Table of Integral Transforms*, Vol. I, Bateman Manuscript Project (McGraw-Hill, 1954).
- [19] Mountain, R. D., *J. Chem. Phys.* **57**, 4346 (1972).